Econometrics

	Course Calendar
Week	Main Content
Week 7	Extension of Simple Regression: Functional Forms I
Week 8	Extension of Simple Regression: Functional Forms II
Week 9	Extension of Simple Regression: Functional Forms III
Week 10	Multiple Regression
Week 11	Multiple Regression: Problem of Inference
Week 12	Multiple Regression: Functional Forms
Week 13	Introduction to Dummy Variables
Week 14	Introduction to Dummy Variables and Regression Methods
Week 15	Regression with Dummy Variables: Hands-on-Exercise
Week 16	Application of Regression

Econometrics

Lecture 12. Multiple Regression Analysis: Inference (continuation) & Functional Forms

Geetha Rani Prakasam, Ph.D

Professor,

Recap

- Multiple Linear Regression: multicollinearity, Partial Regression Coefficients, Adjusted R^2
- Testing of Hypothesis:
- **The Normality Assumption**
- **Hypothesis Testing in MR: General Comments**
  - Testing hy. about an individual partial regression coefficient
  - Testing the overall significance of slope coefficients and the F test

Outline

- **Hypothesis Testing in MR (continuation)**
- The “Incremental” or “Marginal” contribution of an explanatory variable
- Testing the equality of two regression coefficients
- **Restricted least square**
- Comparing two regressions:
- Testing the stability of the estimated regression model over time or in different cross-sectional units
- Testing the functional form of regression

8-6. The “Incremental” or “Marginal” contribution of an explanatory variable:

- Consider the CM example of MR
- Suppose we introduce PGNP and FLR sequentially; i.e., we first regress child mortality on PGNP and assess its significance and then add FLR to the model to find out whether it contributes anything (the order in which PGNP and FLR enter can be reversed).
- By contribution we mean whether the addition of the variable to the model increases ESS (and hence R^2) “significantly” in relation to the RSS.
- This contribution is called the **incremental, or marginal, contribution of an explanatory variable.**

8-6. The “Incremental” or “Marginal” contribution of an explanatory variable:

- Suppose we first regress child mortality on PGNP and obtain the following regression:

$$CM_i = 157.4244 - 0.0114\ PGNP_i \qquad \mathbf{(8.4.14)}$$

$$t = (15.9894)\ (-3.5156)\quad r^2 = 0.1662$$

$$p\ \text{value} = (0.0000)\ (0.0008)\quad \text{adj}\ r^2 = 0.1528$$

- As these results show, PGNP has a significant effect on CM.
- 1. What is the marginal, or incremental, contribution of FLR, knowing that PGNP is already in the model and that it is significantly related to CM?
- 2. Is the incremental contribution of FLR statistically significant?
- 3. What is the criterion for adding variables to the model?

8-6. The “Incremental” or “Marginal” contribution of an explanatory variable:

- These questions can be answered by the ANOVA technique.
- To see this, let us construct the ANOVA Table (next slide).
- In this table X_2 refers to PGNP and X_3 refers to FLR.
- To assess the incremental contribution of X_3 after allowing for the contribution of X_2, we form
- (ESSnew − ESSold)/number of new regressors RSSnew/df (= n − number of parameters in the new model) **(8.4.16)**
- Now applying Eq. (8.4.16), we obtain:
- F = 196,912.9 / 1742.8786 = 112.9814 **(8.4.17)**

8.6. ANOVA Table for the Illustrative Ex.: Incremental Analysis

Source of Variation	SSS	df	MSS
ESS due to PGNP	60,449.5	1	60,449.5
ESS due to the addition of FLR	196,912.9	1	196,912.9
ESS due to PGNP and FLR	257,362.4	2	128,681.2
RSS	106,315.6	61	1742.8786
Total	363,678	63	

R^2 new = 0.7077 (from Eq. [8.1.4]) and
R^2 old = 0.1662 (from Eq. [8.4.14]).
Therefore, F = (0.7077 − 0.1662)/1 (1 − 0.7077)/61 = 113.05 **(8.4.19) ,**
which is about the same as that obtained from Eq. (8.4.17), except for the rounding errors.

8-6. Testing the overall significance of a MR

- <u>Decision Rule:</u>
- If $F_{com} > F_{\alpha, Df1, Df2}$ one can reject the Ho that $\beta = 0$ and conclude that the addition of X to the model significantly increases ESS and hence the R^2 value
- When to Add a New Variable?
- **If $|t|$ of coefficient of X > 1 (or $F= t^2$ of that variable exceeds 1)**
- When to Add a Group of Variables?
- **If adding a group of variables to the model will give F value greater than 1;**

8-7. Testing the equality of two regression coefficients

- $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + u_i$ (8.6.1)
- Test the hypotheses:
- **H_0: $\beta_3 = \beta_4$ or $\beta_3 - \beta_4 = 0$ (8.6.2)**
- **H_1: β_3 # β_4 or $\beta_3 - \beta_4$ # 0**
- Under the classical assumption it can be shown:
- **$t = [(\beta\hat{}_3 - \beta\hat{}_4) – (\beta_3 - \beta_4)] / se(\beta\hat{}_3 - \beta\hat{}_4)$**
- follows the t distribution with (n-4) df because (8.6.1) is a four-variable model or, more generally, with (n-k) df. where k is the total number of parameters estimated, including intercept term.
- **$se(\beta\hat{}_3 - \beta\hat{}_4) = \sqrt{} [var((\beta\hat{}_3) + var(\beta\hat{}_4) – 2cov(\beta\hat{}_3, \beta\hat{}_4)]$ (8.6.4)**

8-7. Testing the equality of two regression coefficients

- $t = (\beta\hat{}_3 - \beta\hat{}_4) / \sqrt{[var((\beta\hat{}_3) + var(\beta\hat{}_4) - 2cov(\beta\hat{}_3, \beta\hat{}_4)]}$ (8.6.5)
- <u>Steps for testing</u>:
- **1. Estimate $\beta\hat{}_3$ and $\beta\hat{}_4$**
- **2. Compute se($\beta\hat{}_3 - \beta\hat{}_4$) through (8.6.4)**
- **3. Obtain t- ratio from (8.6.5) with H_0: $\beta_3 = \beta_4$**
- **4. If t-computed > t-critical at designated level of significance for given df, then reject H_0.**
- Otherwise do not reject it.
- Alternatively, if the p-value of t statistic from (8.6.5) is reasonable low, one can reject H_0.

8-7. Testing the equality of two regression coefficients Ex

- Such a null hypothesis is of practical importance.
- For example, let Eq. (8.6.1) represent the demand function for a commodity where Y = amount of a commodity demanded, X_2 = price of the commodity, X_3 = income of the consumer, and X_4 = wealth of the consumer.
- **The null hypothesis in this case means that the income and wealth coefficients are the same.**
- Or, if Yi and the X's are expressed in logarithmic form, the null hypothesis in Eq. (8.5.2) implies that the income and wealth elasticities of consumption are the same.

EX. 8.2 The Cubic Cost Function

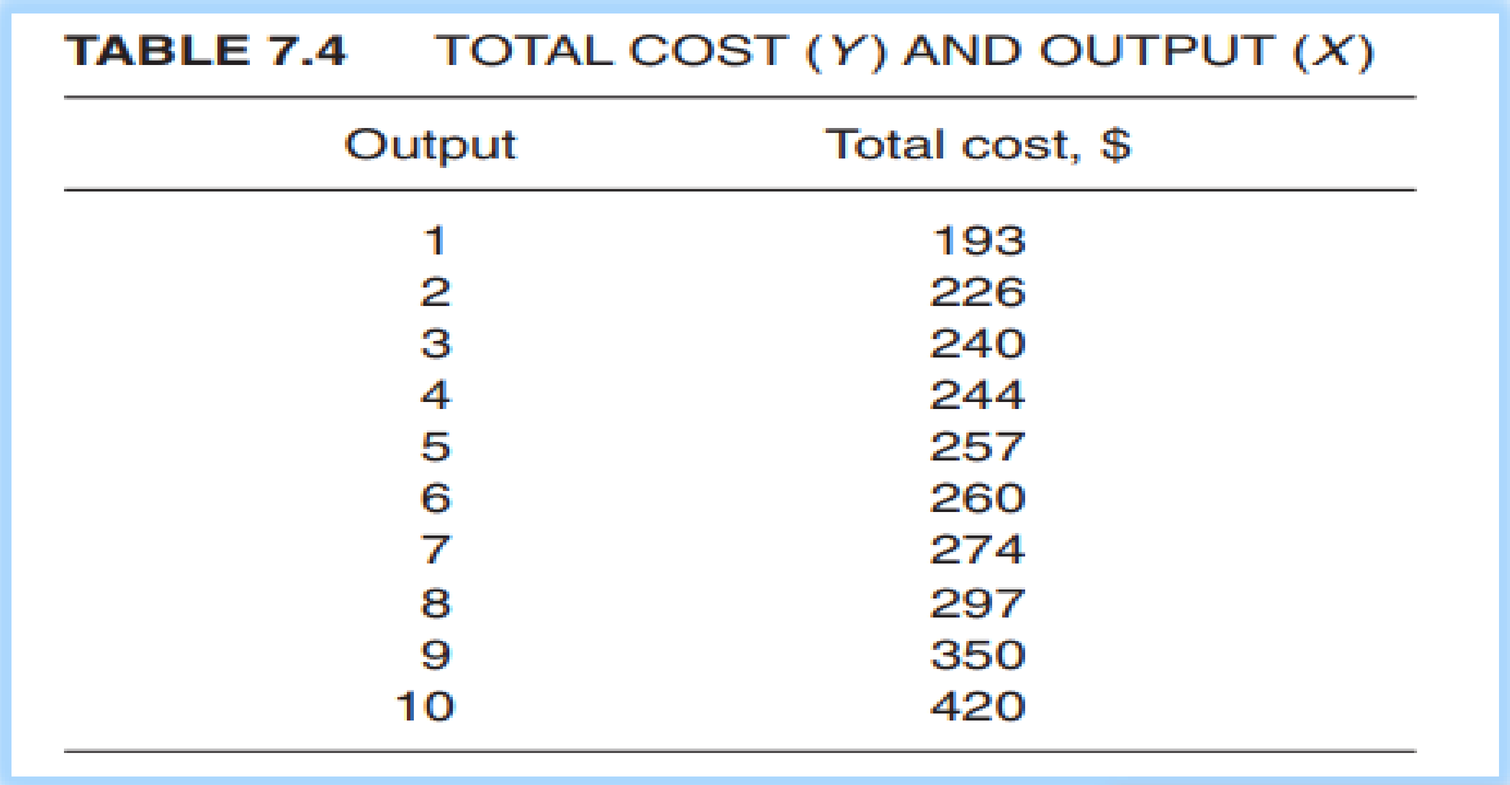
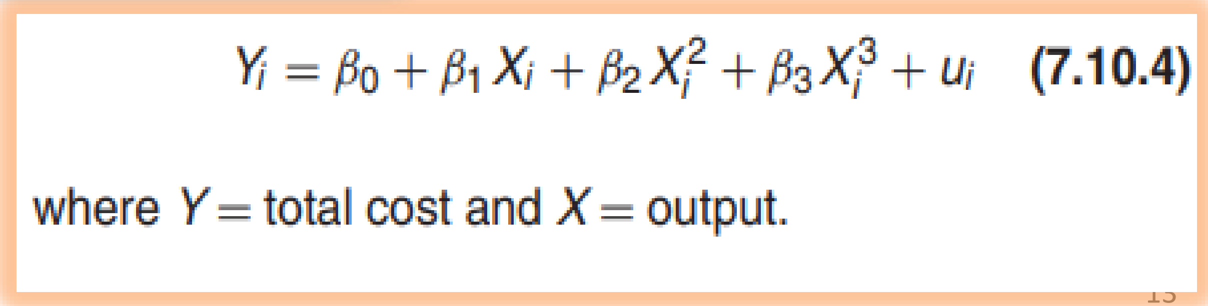


TABLE 7.4 TOTAL COST (Y) AND OUTPUT (X)

Output	Total cost, $
1	193
2	226
3	240
4	244
5	257
6	260
7	274
8	297
9	350
10	420

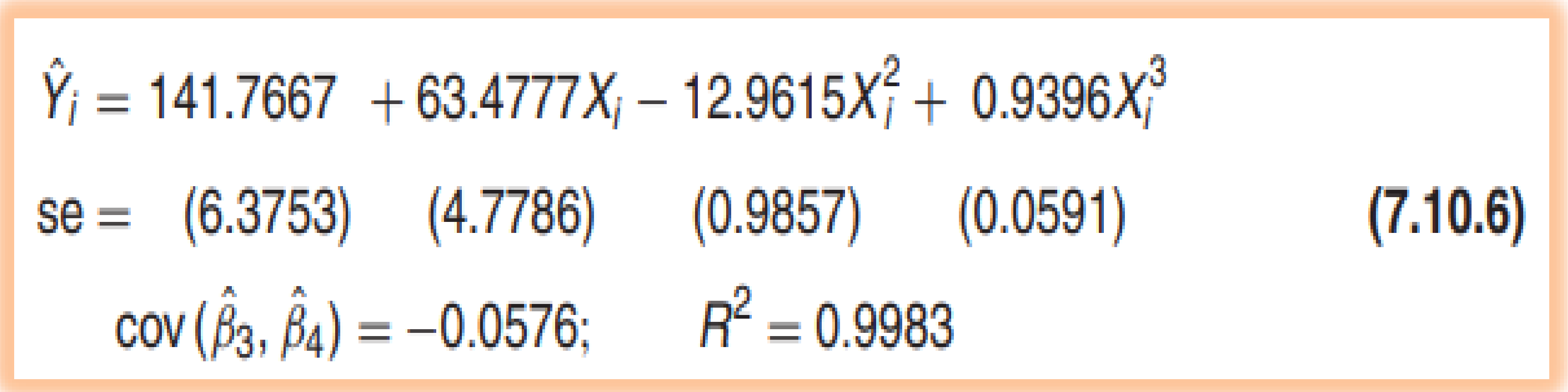


$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + u_i \quad \textbf{(7.10.4)}$$

where Y = total cost and X = output.

EX. 8.2 The Cubic Cost Function

- The cubic total cost function estimated is reported below:



$$\hat{Y}_i = 141.7667 + 63.4777X_i - 12.9615X_i^2 + 0.9396X_i^3$$

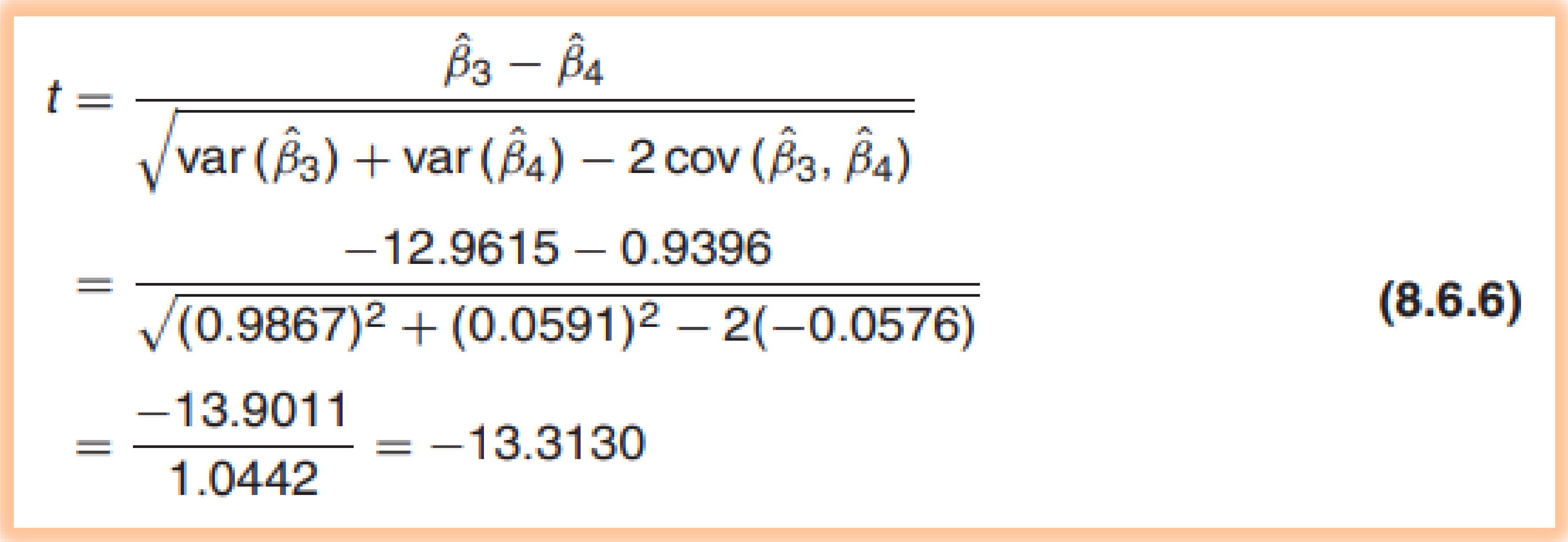
$$\text{se} = \quad (6.3753) \quad (4.7786) \quad (0.9857) \quad (0.0591) \qquad \textbf{(7.10.6)}$$

$$\operatorname{cov}(\hat{\beta}_3, \hat{\beta}_4) = -0.0576; \qquad R^2 = 0.9983$$

- where Y is total cost and X is output, and where the figures in parentheses are the estimated standard errors.
- Suppose we want to test the hypothesis that the coefficients of the X_2 and X_3 terms in the cubic cost function are the same, that is, $\beta_3 = \beta_4$ or $(\beta_3 - \beta_4) = 0$.

EX. 8.2 The Cubic Cost Function

- In the regression (7.10.6) we have all the necessary output to conduct the t test of (8.6.5).
- The actual mechanics are as follows:



$$\begin{aligned} t &= \frac{\hat{\beta}_3 - \hat{\beta}_4}{\sqrt{\operatorname{var}(\hat{\beta}_3) + \operatorname{var}(\hat{\beta}_4) - 2\operatorname{cov}(\hat{\beta}_3, \hat{\beta}_4)}} \\ &= \frac{-12.9615 - 0.9396}{\sqrt{(0.9867)^2 + (0.0591)^2 - 2(-0.0576)}} \qquad \textbf{(8.6.6)} \\ &= \frac{-13.9011}{1.0442} = -13.3130 \end{aligned}$$

EX. 8.2 The Cubic Cost Function

- We can verify that for 6 df (why?) the observed t value exceeds the critical t value even at the 0.002 (or 0.2 percent) level of significance (two-tail test); the p value is extremely small, 0.000006.
- Hence we can reject the null hypothesis that the coefficients of X_2 and X_3 in the cubic cost function are identical.

8-8. Restricted least square:

Testing linear equality restrictions

$Y_i = \beta_1 X_{2i}^{\beta 2} X_{3i}^{\beta 3} e^{u_i}$ (7.10.1) and (8.6.1)

Y = output

X_2 = labor input

X_3 = capital input

In the log-form:

$\ln Y_i = \beta_0 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + u_i$ (8.6.2)

with the constant return to scale: $\beta_2 + \beta_3 = 1$ (8.6.3)

8-8. Restricted least square:

- Testing linear equality restrictions
- How to test the restriction in eq.(8.6.3)?
- The t Test approach (unrestricted): test of the hypothesis H_0: $\beta_2 + \beta_3 = 1$ can be conducted by t- test:
- **$t = [(\beta\hat{}_2 + \beta\hat{}_3) - (\beta_2 + \beta_3)] / se(\beta\hat{}_2 - \beta\hat{}_3)$** **(8.6.4)**
- The F Test approach (restricted least square -RLS):
- Using, say, $\beta_2 = 1-\beta_3$ and substitute it into (8.6.2) we get:
- $\ln(Y_i / X_{2i}) = \beta_0 + \beta_3 \ln(X_{3i} / X_{2i}) + u_i$ (8.6.8).
- Where (Y_i / X_{2i}) is output/labor ratio, and
- (X_{3i} / X_{2i}) is capital/labor ratio

8-8. Restricted least square

- **Testing linear equality restrictions**
- $\Sigma u\hat{}^2_{UR}$=RSS_{UR} of unrestricted regression (8.6.2)
- and Σ $u\hat{}^2_R$ = RSS_R of restricted regression (8.6.7),
- m = number of linear restrictions;
- k = number of parameters in the unrestricted regression;
- n = number of observations.
- R^2_{UR} and R^2_R are R^2 values obtained from unrestricted and restricted regressions respectively.

8-8. Restricted least square:

- Then, $F=[(RSS_R - RSS_{UR})/m]/[RSS_{UR}/(n-k)] =$
- $= [(R^2_{UR} - R^2_R) / m] / [1 - R^2_{UR} / (n-k)]$ (8.6.10)
- follows F distribution with m, (n-k) df.
- **Decision rule: If $F > F_{m, n-k}$, reject H_0: $\beta_2 + \beta_3 = 1$**
- **Testing linear equality restrictions**
- Note: $R^2_{UR} \geq R^2_R$ (8.6.11)
- and $\Sigma u\hat{}2_{UR} \leq \Sigma u\hat{}2_R$ (8.6.12)
- Ex: 8.3: The Cobb– Douglas Production Function for the Mexican Economy, 1955–1974

TAB 8.8: Real GDP, Employment, &Real Fixed Capital in Mexico

Year	GDP*	Employment†	Fixed capital‡
1955	114043	8310	182113
1956	120410	8529	193749
1957	129187	8738	205192
1958	134705	8952	215130
1959	139960	9171	225021
1960	150511	9569	237026
1961	157897	9527	248897
1962	165286	9662	260661
1963	178491	10334	275466
1964	199457	10981	295378

TAB 8.8: Real GDP, Employment, &Real Fixed Capital in Mexico

Year	GDP*	Employment†	Fixed capital‡
1965	212323	11746	315715
1966	226977	11521	337642
1967	241194	11540	363599
1968	260881	12066	391847
1969	277498	12297	422382
1970	296530	12955	455049
1971	306712	13338	484677
1972	329030	13738	520553
1973	354057	15924	561531
1974	374977	14154	609825

*Millions of 1960 pesos;

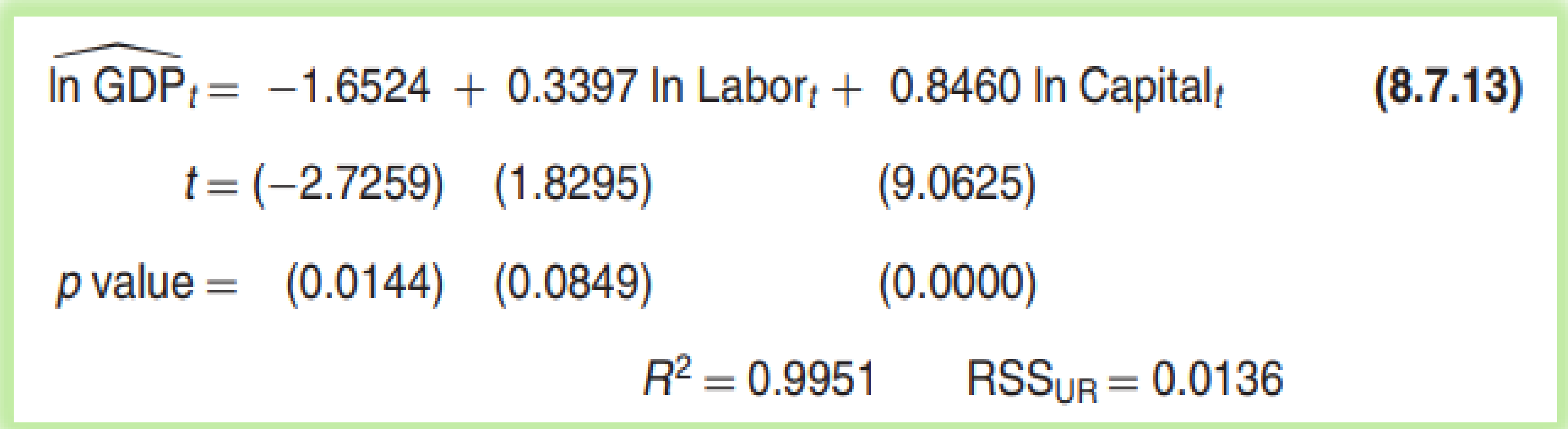
†Thousands of people;
‡Millions of 1960 pesos.

Source: Victor J. Elias, Sources of Growth: A Study of Seven Latin American Economies, International Center for Economic Growth, ICS Press, San Francisco, 1992. Data from Tables E5, E12, and E14.

Source: Basic Econometrics, Damodar Gujarati, Page.270

Ex: 8.3: The Cobb– Douglas Production Function for the Mexican Economy, 1955–1974

- Consider the data given in Table 8.8.
- Attempting to fit the Cobb–Douglas production function to these data, yielded the following results:



$$\widehat{\ln \text{GDP}_t} = -1.6524 + 0.3397 \ln \text{Labor}_t + 0.8460 \ln \text{Capital}_t \qquad \textbf{(8.7.13)}$$

$$t = (-2.7259) \quad (1.8295) \quad (9.0625)$$

$$p \text{ value} = (0.0144) \quad (0.0849) \quad (0.0000)$$

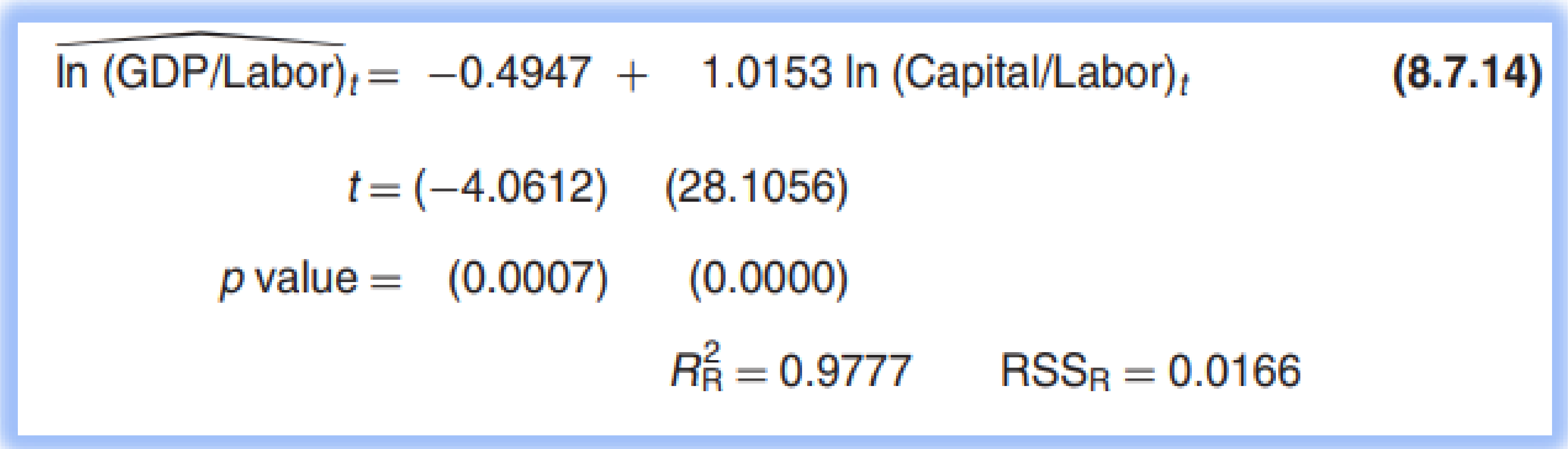
$$R^2 = 0.9951 \qquad \text{RSS}_{\text{UR}} = 0.0136$$

- where RSS_{UR} is the unrestricted RSS, as we have put no restrictions on estimating (8.7.13).

Ex: 8.3: The Cobb– Douglas Production Function for the Mexican Economy, 1955–1974

- To interpret the coefficients of the Cobb– Douglas production function;
- The output/labor elasticity is about 0.34 and the output/capital elasticity is about 0.85.
- If we add these coefficients, we obtain 1.19, suggesting that perhaps the Mexican economy during the stated time period was experiencing increasing returns to scale.
- Of course, we do not know if 1.19 is statistically different from 1.
- To see if that is the case, let us impose the restriction of constant returns to scale, which gives the following regression:

Ex: 8.3: The Cobb– Douglas Production Function for the Mexican Economy, 1955–1974



$$\widehat{\ln(\text{GDP/Labor})}_t = -0.4947 + 1.0153 \ln(\text{Capital/Labor})_t \qquad \textbf{(8.7.14)}$$

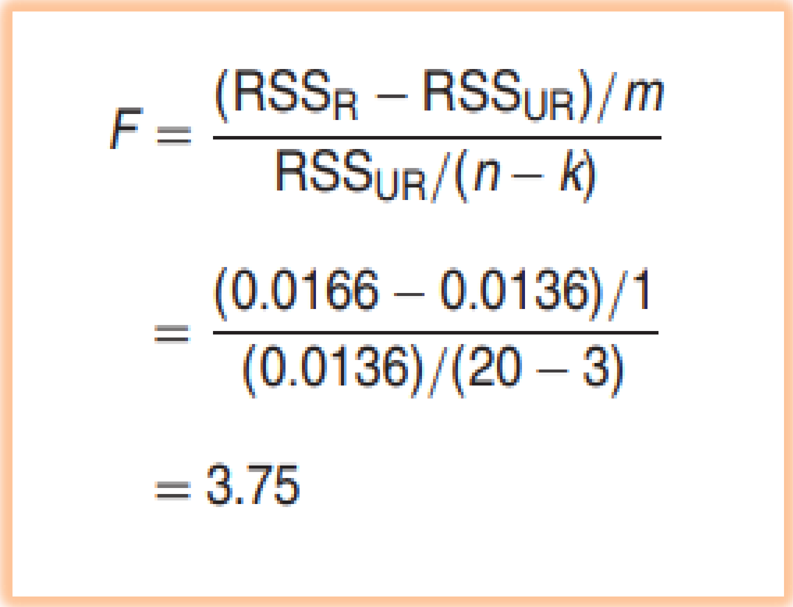
$$t = (-4.0612) \quad (28.1056)$$

$$p\text{ value} = (0.0007) \quad (0.0000)$$

$$R^2_{\text{R}} = 0.9777 \qquad \text{RSS}_{\text{R}} = 0.0166$$

- where RSSR is the restricted RSS, for we have imposed the restriction that there are constant returns to scale.
- Since the dependent variable in the preceding two regressions is different, we have to use the F test given in (8.7.9).
- We have the necessary data to obtain the F value.

Ex: 8.3: The Cobb– Douglas Production Function for the Mexican Economy, 1955–1974



$$F = \frac{(\text{RSS}_\text{R} - \text{RSS}_\text{UR})/m}{\text{RSS}_\text{UR}/(n-k)}$$

$$= \frac{(0.0166 - 0.0136)/1}{(0.0136)/(20-3)}$$

$$= 3.75$$

- Note in the present case m = 1, as we have imposed only one restriction and (n − k) is 17, since we have 20 observations and three parameters in the unrestricted regression.

Ex: 8.3: The Cobb– Douglas Production Function for the Mexican Economy, 1955–1974

- This F value follows the F distribution with 1 df in the numerator and 17 df in the denominator.
- We can easily check that this F value is not significant at the 5% level.
- The conclusion then is that the Mexican economy was probably characterized by constant returns to scale over the sample period and therefore there may be no harm in using the restricted regression given in (8.7.14).
- As this regression shows, if capital/labor ratio increased by 1 percent, on average, labor productivity went up by about 1 percent.

8-9. Comparing two regressions:

- **Testing for Structural or Parameter Stability of Regression Models: The Chow Test**

- When we use a regression model involving time series data, it may happen that there is a **structural change in the relationship between the regressand *Y and the regressors.***

- By structural change, we mean that the values of the parameters of the model do not remain the same through the entire time period.

8.9.Testing for Structural or Parameter Stability of Regression Models:

- Sometime the structural change may be due to external forces (e.g., the oil embargoes imposed by the OPEC oil cartel in 1973 and 1979 or the Gulf War of 1990–1991),
- or due to policy changes (such as the switch from a fixed exchange-rate system to a flexible exchange-rate system around 1973)
- or action taken by Congress (e.g., the tax changes initiated by President Reagan in his two terms in office or changes in the minimum wage rate) or to a variety of other causes.
- Ex policy change or program change Macro Economic Reforms in India since 1991, etc
- During the pandemic period of covid-19 – 2020-2021

8.9.Testing for Structural or Parameter Stability of Regression Models:

- How do we find out that a structural change has in fact occurred?
- Table 8.9: Savings and Personal Disposable Income (billions of dollars), United States, 1970–1995
- Savings function:
- Reconstruction period:
- $Y_t = \alpha_1 + \alpha_2 X_t + U_{1t}$ **(t = 1,2,...,n$_1$)**
- Post-Reconstruction period:
- $Y_t = \beta_1 + \beta_2 X_t + U_{2t}$ **(t = 1,2,...,n$_2$)**

Table 8.9: Savings &Personal Disposable Income (billions $), US, 1970–1995

Observation	Savings	Income	Observation	Savings	Income
1970	61	727.1	1983	167	2522.4
1971	68.6	790.2	1984	235.7	2810
1972	63.6	855.3	1985	206.2	3002
1973	89.6	965	1986	196.5	3187.6
1974	97.6	1054.2	1987	168.4	3363.1
1975	104.4	1159.2	1988	189.1	3640.8
1976	96.4	1273	1989	187.8	3894.5
1977	92.5	1401.4	1990	208.7	4166.8
1978	112.6	1580.1	1991	246.4	4343.7
1979	130.1	1769.5	1992	272.6	4613.7
1980	161.8	1973.3	1993	214.4	4790.2
1981	199.1	2200.2	1994	189.4	5021.7
1982	205.5	2347.3	1995	249.3	5320.8

Source: Economic Report of the President, 1997, Table B-28, p. 332

Basic Econometrics, Damodar Gujarati, Page. 274

8.9.Testing for Structural or Parameter Stability of Regression Models:

- This table gives data on disposable personal income and personal savings, in billions of dollars, for the United States for the period 1970–1995.
- Suppose we want to estimate a simple savings function that relates savings (Y) to disposable personal income DPI (X).
- Since we have the data, we can obtain an OLS regression of Y on X.
- But if we do that, we are maintaining that the relationship between savings and DPI has not changed much over the span of 26 years.

8.9.Testing for Structural or Parameter Stability of Regression Models:

- That may be a tall assumption. For example, it is well known that in 1982 the United States suffered its worst peacetime recession.
- The civilian unemployment rate that year reached 9.7 percent, the highest since 1948.
- An event such as this might disturb the relationship between savings and DPI.
- To see if this happened, let us divide our sample data into two time periods: 1970–1981 and 1982–1995, the pre- and post-1982 recession periods.

Ex: Table 8.9: Savings and Personal Disposable Income (billions $), US, 1970–1995

- Now we have three possible regressions:
- Time: 1970–1981: $Y_t = \lambda_1 + \lambda_2 X_t + u_{1t}\ n_1 = 12$ ***(8.8.1)***
- Time: 1982–1995: $Y_t = \gamma_1 + \gamma_2 X_t + u_{2t} n_2 = 14$ ***(8.8.2)***
- Time: 1970–1995: $Y_t = \alpha_1 + \alpha_2 X_t + u_t n = (n_1 + n_2) = 26$ ***(8.8.3)***
- Regression (8.8.3) assumes that there is no difference between the two time periods and therefore estimates the relationship between savings and DPI for the entire time period consisting of 26 observations.
- In other words, this regression assumes that the intercept as well as the slope coefficient remains the same over the entire period; that is, there is no structural change.

Ex: Table 8.9: Savings and Personal Disposable Income (billions $), US, 1970–1995

- If this is in fact the situation, then $\alpha_1 = \lambda_1 = \gamma_1$ and $\alpha_2 = \lambda_2 = \gamma_2$.
- Regressions (8.8.1) and (8.8.2) assume that the regressions in the two time periods are different; that is, the intercept and the slope coefficients are different, as indicated by the subscripted parameters.
- In the preceding regressions, the u's represent the error terms and the n's represent the number of observations.
- For the data given in Table 8.9, the empirical counterparts of the preceding three regressions are as follows:

Ex: Table 8.9: Savings and Personal Disposable Income (billions $), US, 1970–1995

- $\hat{Y}_t = 1.0161 + 0.0803\ X_t$

  $t = (0.0873)\ (9.6015)$ (8.8.1a)

  $R^2 = 0.9021;\ RSS_1 = 1785.032\ df = 10$
- $\hat{Y}_t = 153.4947 + 0.0148X_t$

  $t = (4.6922)\ (1.7707)$ (8.8.2a)

  $R^2 = 0.2971\ RSS_2 = 10{,}005.22;\ \ df = 12$
- $\hat{Y}_t = 62.4226 + 0.0376\ X_t + \cdots$

  $t = (4.8917)\ (8.8937) + \cdots$ (8.8.3a)

  $R^2 = 0.7672\ RSS_3 = 23{,}248.30\ df = 24$

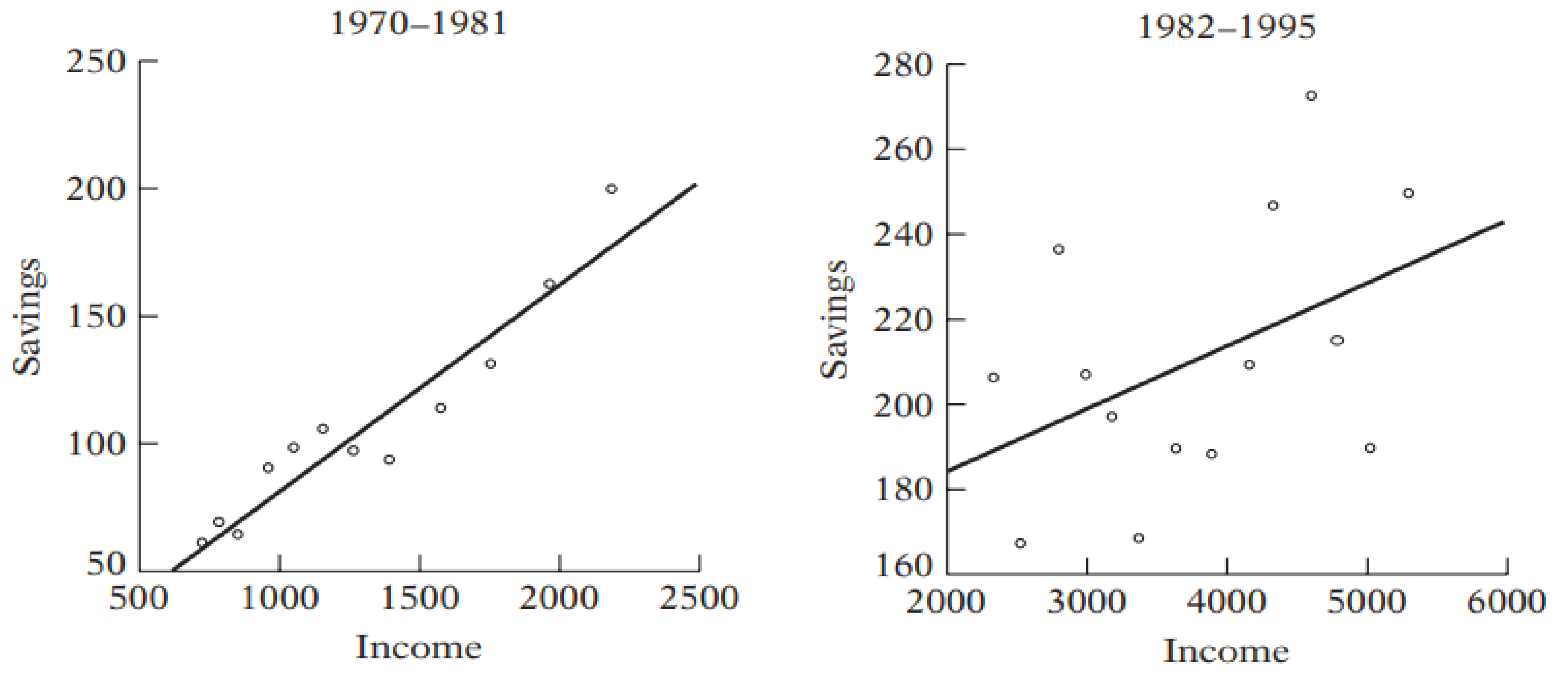
Ex: Table 8.9: Savings and Personal Disposable Income (billions $), US, 1970–1995

- In the preceding regressions, RSS denotes the residual sum of squares, and the figures in parentheses are the estimated t values.
- A look at the estimated regressions suggests that the relationship between savings and DPI is not the same in the two subperiods.
- The slope in the preceding savings-income regressions represents the marginal propensity to save (MPS), that is, the (mean) change in savings as a result of a dollar’s increase in disposable personal income.
- In the period 1970–1981 the MPS was about 0.08, whereas in the period 1982–1995 it was about 0.02.

Ex: Table 8.9: Savings and Personal Disposable Income (billions $), US, 1970–1995

- Whether this change was due to the economic policies pursued by President Reagan is hard to say.
- This further suggests that the pooled regression (8.8.3a)—that is, the one that pools all the 26 observations and runs a common regression, disregarding possible differences in the two subperiods may not be appropriate.
- Of course, the preceding statements need to be supported by appropriate statistical test(s).
- Incidentally, the scattergrams and the estimated regression lines are as shown in Figure 8.3.

FIGURE 8.3



Source: Basic Econometrics, Damodar Gujarati, Page.276

Ex: Table 8.9: Savings and Personal Disposable Income (billions of dollars), US, 1970–1995

- Now the possible differences, that is, structural changes, may be caused by differences in the intercept or the slope coefficient or both.
- How do we find that out?
- A visual feeling about this can be obtained as shown in Figure 8.3. But it would be useful to have a formal test.
- **Chow test:** This test assumes that:
- 1. $u_{\mathbf{1t}} \sim N(0, \sigma\ 2)$ and $u_{\mathbf{2t}} \sim N(0, \sigma\ 2)$.
- That is, the error terms in the subperiod regressions are normally distributed with the same (homoscedastic) variance $\sigma^{2.}$
- 2. The two error terms $u_{\mathbf{1t}}$ and $u_{\mathbf{2t}}$ are independently distributed.

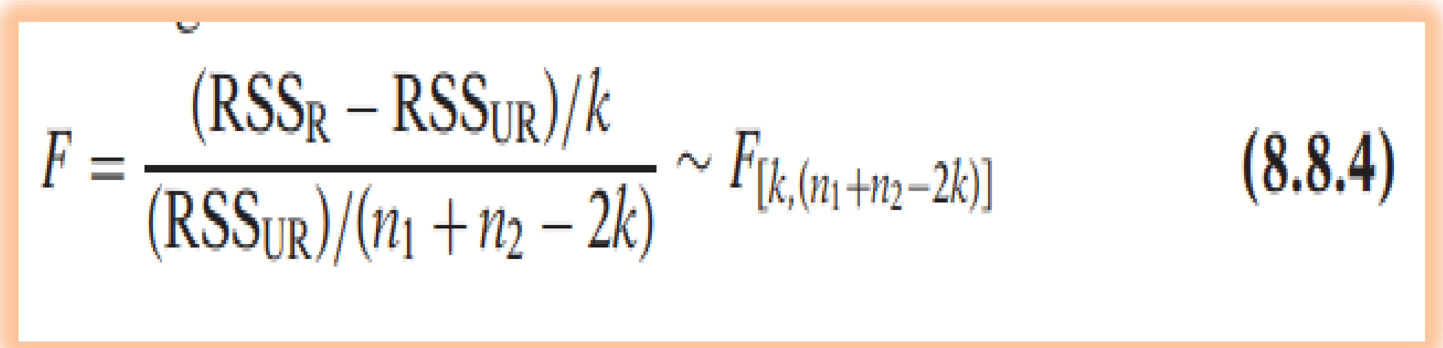
8.9.The mechanics of the Chow test are as follows:

- 1. Estimate regression (8.8.3), which is appropriate if there is no parameter instability, and obtain RSS_3 with df = $(n_1 + n_2 - k)$, where k is the number of parameters estimated, 2 in the present case.
- For our example RSS_3 = 23,248.30. We call RSS_3 the restricted residual sum of squares (RSS_R) because it is obtained by imposing the restrictions that $\lambda_1 = \gamma_1$ and $\lambda_2 = \gamma_2$, that is, the subperiod regressions are not different.
- 2. Estimate (8.8.1) and obtain its residual sum of squares, RSS1, with df = $(n_1 - k)$. In our example, RSS_1 = 1785.032 and df = 10.

8.9.The mechanics of the Chow test are as follows:

- 3. Estimate (8.8.2) and obtain its residual sum of squares, RSS_2, with df = (n_2 − k). In our example, RSS_2 = 10,005.22 with df = 12.
- 4. Since the two sets of samples are deemed independent, we can add RSS_1 and RSS_2 to obtain what may be called the unrestricted residual sum of squares (RSS_{UR}), that is, obtain:
- $RSS_{UR} = RSS_1 + RSS_2$ with df = $(n_1 + n_2 - 2k)$
- In the present case,
- RSS_{UR} = (1785.032 + 10,005.22) = 11,790.252
- 5. Now the idea behind the Chow test is that if in fact there is no structural change [i.e., regressions (8.8.1) and (8.8.2) are essentially the same], then the RSS_R and RSS_{UR} should not be statistically different. Therefore, if we form the following ratio:

8.9.The mechanics of the Chow test are as follows:



$$F = \frac{(\mathrm{RSS_R} - \mathrm{RSS_{UR}})/k}{(\mathrm{RSS_{UR}})/(n_1 + n_2 - 2k)} \sim F_{[k,(n_1+n_2-2k)]} \qquad \mathbf{(8.8.4)}$$

- then Chow has shown that under the null hypothesis the regressions (8.8.1) and (8.8.2) are (statistically) the same (i.e., no structural change or break) and the F ratio given above follows the F distribution with k and ($n_{\mathbf{1}}$ + $n_{\mathbf{2}}$ − 2k) df in the numerator and denominator, respectively.
- 6. Therefore, we do not reject the null hypothesis of parameter stability (i.e., no structural change) if the computed F value in an application does not exceed the critical F value obtained from the F table at the chosen level of significance (or the p value). In this case we may be justified in using the pooled (restricted?) regression (8.8.3).

8.9.The mechanics of the Chow test

- Contrarily, if the computed F value exceeds the critical F value, we reject the hypothesis of parameter stability and conclude that the regressions (8.8.1) and (8.8.2) are different, in which case the pooled regression (8.8.3) is of dubious value, to say the least.
- Returning to our example, we find that
- $F = [(23{,}248.30 - 11{,}790.252)/2]/\ (11{,}790.252)/22$
- $= 10.69$ (8.8.5)
- From the F tables, we find that for 2 and 22 df the 1 percent critical F value is 5.72.
- Therefore, the probability of obtaining an F value of as much as or greater than 10.69 is much smaller than 1 percent; actually the p value is only 0.00057.

8.9.The mechanics of the Chow test

- The Chow test therefore seems to support our earlier hunch that the savings–income relation has undergone a structural change in the United States over the period 1970–1995, assuming that the assumptions underlying the test are fulfilled.
- We will have more to say about this shortly. Incidentally, note that the Chow test can be easily generalized to handle cases of more than one structural break.
- For example, if we believe that the savings–income relation changed after President Clinton took office in January 1992, we could divide our sample into three periods: 1970–1981, 1982–1991, 1992–1995, and carry out the Chow test.
- Of course, we will have four RSS terms, one for each subperiod and one for the pooled data. But the logic of the test remains the same.

Some caveats about the Chow test

- There are some caveats about the Chow test that must be kept in mind:
- 1. The assumptions underlying the test must be fulfilled. For example, one should find out if the error variances in the regressions (8.8.1) and (8.8.2) are the same. We will discuss this point shortly.
- 2. The Chow test will tell us only if the two regressions (8.8.1) and (8.8.2) are different, without telling us whether the difference is on account of the intercepts, or the slopes, or both. But in Lecture 13, on dummy variables, we will see how we can answer this question.
- 3. The Chow test assumes that we know the point(s) of structural break. In our example, we assumed it to be in 1982. However, if it is not possible to

Some caveats about the Chow test

- To determine when the structural change actually took place, we may have to use other methods.
- Before we leave the Chow test and our savings–income regression, let us examine one of the assumptions underlying the Chow test, namely, that the error variances in the two periods are the same.
- Since we cannot observe the true error variances, we can obtain their estimates from the RSS given in the regressions (8.8.1a) and (8.8.2a), namely,

$$\hat{\sigma}_1^2 = \frac{\text{RSS}_1}{n_1 - 2} = \frac{1785.032}{10} = 178.5032 \qquad \textbf{(8.8.6)}$$

$$\hat{\sigma}_2^2 = \frac{\text{RSS}_2}{n_2 - 2} = \frac{10{,}005.22}{14 - 2} = 833.7683 \qquad \textbf{(8.8.7)}$$

Summary and Conclusions

- Steps of Hypothesis testing: in framing the Null and alternative hypothesis
- Identifying the appropriate test and accordingly the formula
- Estimate the test statistics and compare it with the critical values
- Evaluate whether to reject or do not reject the null hypothesis
- The “Incremental” or “Marginal” contribution of an explanatory variable: using the example of Child Mortality as the regressand

Summary and Conclusions

- Testing the equality of two regression coefficients
- **Restricted least square – Examples of cubic cost function; Cobb-Douglas production function**
- Comparing two regressions:
- Testing the stability of the estimated regression model over time – using the savings and income data of US as an example
- This similar testing of hypothesis can be carried out for different cross-sectional units
- Testing the functional form of regression

References

Basic Econometrics by Damodar Gujarati, **Chapter 8: MULTIPLE REGRESSION ANALYSIS: The Problem of Inference**

Basic Econometrics by Damodar Gujarati, **Chapter 7: MULTIPLE REGRESSION ANALYSIS: The Problem of Estimation**

What Next?

- Multiple Linear Regression: The Problem of Inference – continuation
- MR-Different Functional Forms
- Introduction to Dummy Variables